Project 1 Codebook

1. CSV Files used:

* Beers.csv

|  |  |  |
| --- | --- | --- |
| Column Name | Column Type | Description |
| Name | Factor | Name of the beer |
| Beer\_ID | Integer | Unique identifier of the beer |
| ABV | Numerical | Alcohol by volume of the beer |
| IBU | Integer | International Bitterness Units of the beer |
| Brewery\_ID | Integer | Brewery id associated with the beer |
| Style | Factor | Style of the beer |
| Ounces | Numerical | Ounces of beer |

* Breweries.csv

|  |  |  |
| --- | --- | --- |
| Column Name | Column Type | Description |
| Brew\_ID | Integer | Unique identifier of the brewery |
| Name | Factor | Name of the brewery |
| City | Factor | City where the brewery is located |
| State | Factor | U.S. State where the brewery is located |

1. Dataframes created in the Project1.rmd file

* BeersBreweries:
  + Combination of beers.csv and breweries.csv files. All column names have the same characteristics as above except for as outlined below

|  |  |  |
| --- | --- | --- |
| Column Name | Column Type | Description |
| Brewery\_id | Integer | Combined/Merged Brewery\_id and Brew\_ID columns from beers.csv/breweries.csv data set |

* MissingIBU
  + BeerBreweries datframe filtered for missing IBU values
* MissingABV
  + BeersBreweries dataframe filtered for missing ABV values
* MissingStyle
  + BeersBreweries dataframe filtered for missing Style values
* IBUClean
  + BeerBreweries datframe filtered to exclude missing IBU values
* ABVClean
  + BeerBreweries datframe filtered to exclude missing ABV values
* IBUCleanSummary
  + IBUClean dataframe grouped and summarized by state
* ABVCleanSummary
  + ABVClean dataframe grouped and summarized by state
* IBUClean2
  + IBUClean dataframe sorteed for highest individual IBU value
* ABVClean2
  + ABVClean dataframe sorted for highest individual ABV value
* IBUABVClean
  + ABVClean dataframe filtered to exclude missing IBU values

|  |  |  |
| --- | --- | --- |
| Column Name | Column Type | Description |
| StyleGroup | Character | New column created based on text recognition that creates grouped style labels based on key words in the Style column |

* Fit
  + Linear regression model ran on the IBUABVClean dataframe
* IBUABVCleanScaled
  + IBUABVClean dataframe with scaled values for IBU and ABV
* Classifications
  + KNN model ran on the IBUABVCleanScaled dataframe for k=3
* Classifications2
  + KNN model ran on the IBUABVCleanScaled dataframe for k=7
* Classifications3
  + KNN model ran on the IBUABVCleanScaled dataframe for k=11
* MeanAcc
  + Mean Accuracy for the naive bayes model ran on 1000 iterations
* StyleClean
  + BeerBreweries datframe filtered to exclude missing Style values

|  |  |  |
| --- | --- | --- |
| Column Name | Column Type | Description |
| StyleGroup | Character | New column created based on text recognition that creates grouped style labels based on key words in the Style column |

* theText
  + Words in the Style column of the StyleClean dataframe are unlisted and separated to individual
* FindMoreTypes
  + Words in the Style column of the StyleClean dataframe are unlisted and separated to individual
* TotalCount
  + StyleClean column grouped by State
* TotalCountbyStatePer
  + Total count column grouped by State with mutated column for percentages
* Us\_states
  + US state coordinates data

|  |  |  |
| --- | --- | --- |
| Column Name | Column Type | Description |
| long | Numerical | Longitude values by state |
| lat | Numerical | Latitude values by state |
| group | Numerical | State group values |
| order | Integer | Ordering |
| region | Character | Names of the States |
| subregion | Character | Names of state regions (not used) |

* State\_abbs

|  |  |  |
| --- | --- | --- |
| Column Name | Column Type | Description |
| state | Character | State names |
| abb | Character | Abbreviation for state names |

* TopTotalCountbyStatePer2
  + Merged the TotalCountbyStatePer dataframe with the state\_abbs dataframe
* TotalCountbyStylePer
  + Total count dataframe grouped by StyleGroup
* TotalCountbyStylePerAle
  + TotalCountbyStylerPer dataframe filtered for Stylegroup = “Ale”
* TotalCountbyStylePerIPA
  + TotalCountbyStylerPer dataframe filtered for Stylegroup = “IPA”